Third Way for AI Development: Introducing the Architecture of Even Distribution of Synthetic Intelligence in Society [1]

Perspective Article

Author: Damian Górski, contact@damiangorski.com

Keywords: Artificial Intelligence, Existential Risk, AI Safety, Democracy, AI Race, Sustainability, Global Policy

Abstract

The rapid development of Artificial Intelligence creates both opportunities and risks. Although AI has the potential to solve many issues facing humanity, the current models of societies organization increase the spectrum and scale of threats from this direction. This text presents an alternative approach on how societies can function and develop in the emerging age of AI. The new approach is based on the concepts of even distribution of Synthetic Intelligence in society, personal computing power (PCP) inalienably assigned to each human being, and the architecture of local computing clusters belonging to citizens and operating in a common, standardized technology stack to provide unified environment for deploying AI systems. The new proposal is, in fact, the restoration and strengthening of the best-proven approach on our planet provided over millennia by biological evolution: a relatively even distribution of intelligence within the entire species. The first section of the article focuses on determining main AI threats amplified by current models of societies organization (the threat of so-called "race to the precipice" that may lead entire humanity toward a catastrophe as well as threat of concentration of power by narrow group as a result of leveraging the significant AI potential). The following sections outline the key concepts of the new model, a set of social and technological changes that should be implemented, scenarios of international propagation, and possible impact on the de-escalation of AI threats.

1. Main AI threats and current models of societies functioning

Over the past few years, progress in the field of deep learning has led to the advent of intelligent systems with impressive capabilities.[2, 3, 4] With each passing month, AI is becoming a more advanced and universally applied technology possessing the potential to completely transform our world.[5, 6] Although it can help solve many pressing problems, it can also pose a serious threat to the safety of individuals, the stability of democracy, and even the survival of *Homo sapiens.*[7, 8, 9] Particularly worrisome is that the current effort being directed towards AI safety is significantly lower than those being placed into creating more and more powerful systems of this type.[10] It is estimated that currently, safety area constitutes less than 1% of the total expenditure on AI research and development.[11] The rush to develop and deploy these technologies as fast as possible is currently driven by two types of entities: companies competing within the capitalist system and, from a broader perspective, by governments competing on the international stage. In the reality of the conflict of interest between competing companies and entire countries, as well as conflict of core values between global powers (especially China and the US), entities are particularly motivated to retain and strengthen their status.[12, 13, 14] In such an environment, developing powerful AI systems as quickly as possible and at a large scale can provide significant advantages over competitors. At the same time, for entities that fall behind in this race, their status can become more and more degraded and more difficult to restore. To gain an advantageous position, companies and countries need significant computing power and high-quality know-how that are essential to develop more and more powerful Language Learning Models (LLMs).[15, 16, 17] Accumulated know-how and computing power can provide substantial advantages over

competitors and, consequently, help further strengthen their status (concentration of capital, resources, and authority).

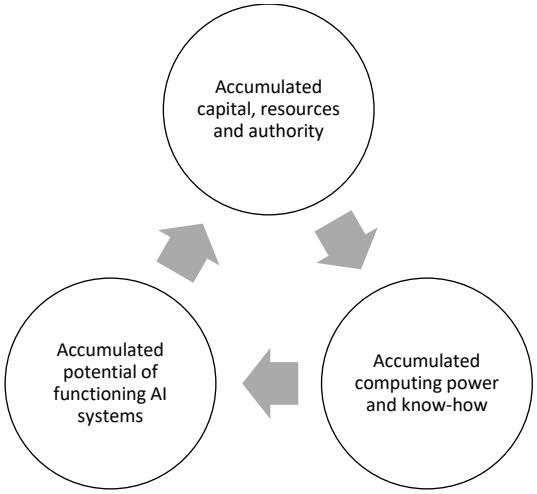

Fig. 1: A diagram of a Synthetic Intelligence consolidation cycle over time.

There are two main types of risks related to the current state of AI development, which will be described below. The first one concerns the risk of a catastrophe that can degrade the status or even eliminate the entire *Homo sapiens* species as a result of the reckless and unsustainable development of AI. The second one concerns the risk of consolidated AI potential by large entities (such as governments and/or corporations), thereby leading to a highly detrimental one-sided dependency relationship for the underprivileged part of society.

Main Risk 1: Regional and global "race to the precipice" leading to a catastrophe for humanity.

The strong desire of companies and governments to leverage the potential of AI as quickly and extensively as possible in order to strengthen their status and, on the other hand, their fear of degradation due to the actions of competing actors may drive an increasingly risky race in which safety is not the top priority. Moreover, entities may also be motivated to continue rapid progress of AI by fears that other actors will do so in a more reckless and potentially dangerous way for mankind. The reasoning may be as follows: if I don't race to create more and more powerful AI, another entity, possibly one that is more selfish and unpredictable, will. The actions that companies and governments take on the basis of strong pressure and fear do not favor building social trust, transparency, caution, diligence, and, ultimately sustainability and safety of AI. This state of affairs can lead to more and more ruthless and risky race to the precipice where finally, nobody from our species will be on the winning side.[18, 19] The continuation of the race to the precipice may, among other things, lead to the following negative consequences for humanity:

a) Loss of control over powerful AI – losing control over increasingly powerful AI. [20, 21, 22] Such "liberated" AI can become completely independent from humanity. Its values, goals, and actions can be radically different from ours at an early stage or its alignment may decrease over time. Losing control over powerful AI may, among others, lead to the enslavement or complete annihilation of our species, which may be perceived as unimportant, useless, or harmful.

b) Catastrophic incidents caused by the misuse of AI or its derivative technologies by humans – the potential of AI and its derivative technologies may be targeted at any area, including potentially

risky areas of chemical technologies, bioengineering, and nanotechnologies.[23, 24] Moreover, reckless AI progress driven by pressure and fear can intensify local and global social tensions, leading to conflicts.[25, 26] The above trends could increase the risk of the creation of dangerous incidents that could result in radioactive contamination, a pandemic with a high mortality rate, or the release of life-threatening devices to the environment based on nanotechnologies.

c) Common social issues of mental nature – careless, socially unbalanced AI development may lead to deep and common disorders among society, such as interpersonal relations disorders, losing one's self-worth, lack of motivation and purpose (e.g., caused by automation and technological unemployment). [27, 28] In the long term, this can negatively impact the functioning of societies, leading to a surge in suicides or a decline in fertility rates and consequently, population collapse.

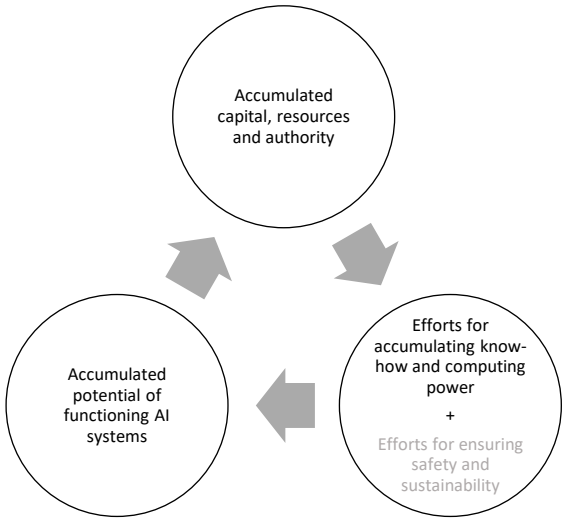

Fig. 2: Expenditures on safe and socially responsible AI development may be a significant financial cost for entities. Additionally, it may extend the time of developing systems ready to be implemented. The entities which do not attach much importance to these crucial aspects while having competitive R&D potential and computing power can be ahead of the entities which are more socially responsible.

Main Risk 2: Concentration of power by entities that consolidated significant AI potential. The decrease of safety of the rest of society caused by strong disproportion of statuses and by transition from two-sided interdependence relationship of both groups to one-sided dependence.

The accumulation of a significant amount of computing power and AI know-how by a single entity or a narrow group of entities can widen the status disparity between them and the rest of society.[29] Leveraging their privileged position, such actors may be able to use an ever-widening range of digital technologies as well as techniques of social engineering, consequently expanding their ability to influence the views, behaviors and actions of the rest of society, leading to their increasing dominance over time.[30] Such tendencies can deeply affect the functioning of free societies and democracies, leading to their erosion and a gradual transition to autocracy.[31] Despite these unfavorable implications, as long as the group that consolidated AI potential is dependent on the rest of society, the society has a fairly high chance of retaining its existence (although its conditions and overall quality remain in question in such a situation). Such state of affairs is due to the presence of a two-sided (mutual) dependence relationship of both groups. Throughout the entire history of societies, the status of every individual was largely determined by the potential of the remaining members of a given community. Moreover, the success of individuals who have particularly wide capabilities of influencing the group, including all kinds of elites, has been always dependent on the remaining members of the society. In

fact, it is the intellectual potential, skills, knowledge, and work of entire society that has determined the power and status of the elites. Over the centuries, this dependency has concerned all privileged groups and individuals, regardless of the times or the social order existing in a given area.

However, in the reality of powerful AI systems, this mutual dependence that has been an uninterrupted part of human history may be irretrievably lost. The entities that have gathered enough computing power and know-how (especially governments and large corporations with adequate resources) can generate efficient intellectual workforce that is not based on biological structures, and instead, relies on silicon structures. Such a situation means the emergence of a substitute for the intellectual human capital, completely changing the 'rules of the game' that have so far existed on our planet. In the reality of consolidated AI potential, intellectual as well as physical capabilities (resulting from AI-driven robotics and automation) cease to be highly limited resources that are only assigned to biological organisms and are fairly evenly distributed across the whole society. In these new circumstances in the history of humankind, the relationship between the entities which will be able to accumulate Synthetic Intelligence potential and the rest of the society may become increasingly weaker and more disadvantageous for the latter. As a result, the relationship between both groups may change from mutual dependence (which, of course, was more or less just and balanced over the ages) to a one-sided relationship of a master–slave nature.

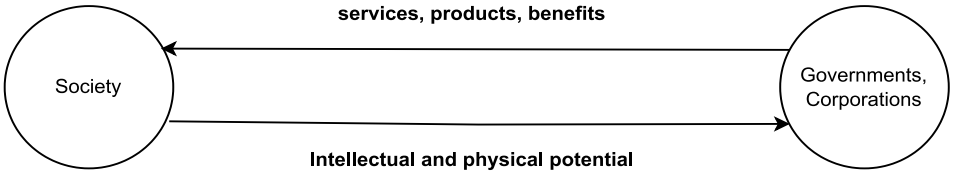

Fig. 3: Human intelligence era and mutual-dependence.

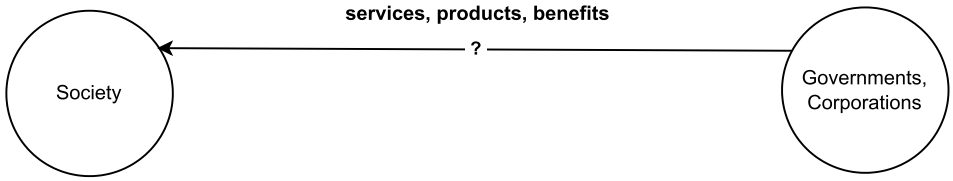

Fig. 4: Synthetic Intelligence era and one-sided dependence.

Importantly, even when assuming that there is intensive and fruitful cooperation between both groups at a given stage, the entities that consolidate significant computing power and know-how are the gatekeepers to the AI potential. Over time, when society begins to use AI systems more broadly and intensively, it becomes more dependent on their providers. On the other hand, the providers, who own more and more AI potential when compared to rest of society, becomes less dependent on it. This state of affairs may lead to a situation where the entities which have accumulated significant AI potential become independent from rest of society (their past customers, workers, and voters). Moreover, as the disparity of the status of the privileged group grows, the discrepancy between their initial (e.g., pro-social and humanistic) and current values may become increasingly wider.

The emergence of a one-sided dependence relationship marks the collapse of the existential social contract that has been in place within our species for thousands of years. Human capital ceases to serve as life insurance for human individuals and a guarantee of maintaining their rights. The consolidation of AI capabilities may thus not only lead to a highly unfavorable disparity of statuses, but may also create the risk of a collapse of the bilateral relationship of dependency that is crucial to the preservation of the existence of all human beings.

Present approaches of societies organization and AI development

Currently, individual countries are implementing the following models of societies organization and AI development:

1) The development of AI by the private sector, especially Big Tech companies - In this approach, the development and providing of AI services mostly takes place by private entities that have a large amount of efficient computing power and advanced know-how.[32, 33, 34] Currently, this model is almost completely unregulated. However, some regulations aimed at increasing safety are being considered.[35]
2) The development of AI by governments and strongly affiliated entities – AI is developed and made available by a government and entities that are strongly affiliated with it (national research institutes, corporations with ties to the government) and possess a large and efficient computing power infrastructure and advanced know-how.[36, 37, 38] Because of centralized power, this approach can be strongly regulated by the government at an early stage.[39]

Can we counteract the aforementioned AI risks by continuing with one of the current models? Or rather, both lead to escalating threats?

In the first approach, the private sector, with its adequate know-how and infrastructure, plays the leading role in training and hosting more and more powerful AI. The LLMs developed and suitably adjusted by companies are then made available to society as online services (such as ChatGPT or Google Bard).[40] There are ongoing discussions about implementing governmental regulations to increase the safety of more and more powerful systems. Though the regulations aiming to avoid risky incidents with AI use that endanger all humanity (especially the risk of losing control over powerful AI) are discussed widely (as it is in the interest of all), the regulations aiming to prevent the accumulation of AI potential by Big Tech companies may be contrary to the goals of these entities, as they are focused on increasing their profits. AI consolidation can provide them an exponential advantage over other companies on both regional and global scales and ultimately, further increase their profits and status. To achieve this as efficiently as possible, corporations that are leading the AI progress can negatively interfere with the society's awareness of AI consolidation's impact on the safety of individuals and the stability of democratic systems, e.g., by conducting misinformation campaigns or information noise.[41] At the same time, their accumulation of substantial resources can allow them to tighten their ties with the government and lobby for solutions that are favorable to their interests and that help maintain and strengthen the tendency of consolidation.[42, 43] In the capitalistic system, where technological progress is driven by the entities emphasizing increasing profits and gaining as great an advantage as possible over competition, social sustainability of AI technology can be relatively low on the priority list.

In the second approach, the AI potential is managed by the government and entities that strongly dependent on the authority (companies and public institutions). This model is currently being pursued in countries with strong centralized power, such as China. The significant difference of this approach when compared to the first model is that the center of power is moved away from corporations toward the government and this may facilitate the coordination of some actions. This approach may also provide hope that profit maximization is not the main driver of the AI progress. At least in theory, the government can be — but absolutely does not have to be — oriented toward its citizens. In an optimistic scenario, this approach can limit the race to the precipice between the corporations, whose actions are curbed by the centralized authority of the government. However, even in such a case, AI consolidation by the government of such a country may negatively impact relations on the international stage. Other countries may be reasonably averse to the global expansion of this autocratic model based on a government's consolidation of powerful AI. This state of affairs may, in turn, lead to the escalation

of international tensions and an even greater race to the precipice between countries on a global level. It should be strongly emphasized that the consolidation of AI potential by any small group does not assure the security of individuals in society. In such a case, the highly cumulative potential of artificial intelligence and power leads to a highly unfavorable one-way dependency relationship for society. Additionally, an entity that accumulates Synthetic Intelligence continues to gain more intelligence and knowledge over time. As a result, its goals and priorities may become less aligned with the goals and interests of the whole society. Importantly, the shift from its original position, covering values, goals, and activities, toward indifference or even aversion may increase as the status gap between both groups widens.

It is also worth being cautious when approaching voices stating that the fastest possible development of AI is best for everyone because this may bring quicker solutions to the pressing issues facing humanity, such as hunger, illness, and poverty. However, it is essential to note that if all the current problems of humanity can be solved by a powerful entity that is not strongly dependent on the entire society, it will not improve our security and chances for further, prosperous existence. If all adverse environmental factors can be minimized by a new, powerful factor (i.e., by a powerful entity driven by AI) that is out of our control and has a much greater impact on our existence than all previous factors, our safety does not increase, but radically decreases.

Summary

The current models of AI development do not ensure the even distribution of Synthetic Intelligence in society. These approaches are highly risky for the underprivileged groups of humanity, while also posing an existential threat to humanity as a whole through a progressive race to the precipice in which Homo sapiens may be on the losing side. This highly unfavorable state of affairs seems to have its roots at the very foundations on which the current models are built. Therefore, it is necessary to create a new model for the functioning of societies and further develop AI within its bounds; a model that will be as resilient to current risks as possible. This is the primary task of this text, which presents a proposal for such a new model.

2. Foundation of the new model

Powerful AI systems require efficient infrastructure to train them and make them available to the public on a large scale. The consolidation of AI potential by entities leads to the accumulation of capital, resources and real power which in itself is a risk at least for the underprivileged part of society. Moreover, companies and government, motivated by the desire to build advantages through AI and by fear of their status being degraded by competitors, are driving a race to the precipice that is not conducive to ensuring the safety of deployed technologies. Unfortunately, this may ultimately lead to a disaster in which there will be no winners on the humanity side.

The above, highly risky tendencies for societies taking place thanks to the positive perception of the digital era innovations over the past few decades which paradoxically widely opened the way for current trends. With the advent of the first personal computers and the development of the fast global decentralized network of the Internet, computer technologies have begun to be viewed as a new, promising, and independent medium. They allowed billions of people to gain skills and knowledge, exchange information on a global scale, and empower freedom of individuals.[44, 45, 46] People have started viewing digital technologies as positively as the earlier, highly emancipating technologies of everyday life, like telegraph, telephone, or car. Because of their positive impact, they have gained the

same social and legal status by default. As a result, just like every person and company currently has the right to own any amount of phones or cars, similarly every entity has the right to own any amount of advanced and efficient CPUs and any amount of computing power necessary to develop powerful AI systems.

Today, however, perceptions of digital technologies in transforming our world and shaping our future are changing from the hopes placed in them in recent decades. Currently, it is not a globally decentralized network of computers but a consolidated power of cloud computing owned by large entities starts determines the proportions of power within society. The computing power of CPUs matches and also surpasses the capabilities of human brains in more areas previously thought to be reserved exclusively for *Homo sapiens*. In effect, computing power becomes a crucial resource for our safety. In this regard, the infrastructure of the computing power that is the basis of the AI revolution should not be longer treated the way society has treated it so far.

Urgent need for fundamental social changes

Free societies have always jointly established rules for the organization of their lives, by continuously adapting the law and order to civilizational challenges. Over the past decades, societies have together decided which technologies are critical for their future and therefore should be treated in a specific way. The legal changes and the implementation of effective safety measures allowed them to minimize the risks following from, , among other things, the civilian and military development and application of potentially destructive nuclear, chemical, and bioengineering technologies.[47, 48, 49]

None of the technologies to date have been as universal and powerful as AI. The advent of non-biological Intelligence and its consolidation by one entity or a narrow group becomes a critical threat to the democratic order and the safety of individuals, including their very existence. In view of the above, free societies should cease tolerating the increasingly risky consolidation tendencies. Instead, computing power and AI potential which is its derivative should be distributed evenly throughout the society and inalienably assigned to each human being. This fundamental rule should be applied to legislation of all the countries that want to ensure safety of every individual in the age of AI. Such an approach is, in fact, a restoration and reinforcement of the best proven strategy on the planet provided over millennia by biological evolution: a relatively even distribution of intelligence within the species.

Possible implications of changes for safety and social stability

Even distribution of Synthetic Intelligence in society means that corporations and governments cannot own computing power and, as a result, cannot consolidate the AI potential that could lead to the degradation of social order and the status of an individual. New approach inextricably binds these entities with society, which becomes the provider of computing power that is necessary to pursue common goals. Instead of consolidation and growing one-sided dependence, mutual dependence is established, which is safer for individuals and social stability.

Importantly, the new model supports developing and implementing next generations of AI systems by corporations or governmental bodies. However, this must be done in line with the will and interests of society, which in new model is the provider of computing power. What's important, the new approach allows combine the computing power of individuals to ensure efficient infrastructure that is necessary for the operation of powerful AI. This possibility may be particularly important in the case of an unfavorable international scenario in which country with significant computing power potential,

despite the alternative based on cooperation under the new model, chooses the path of further narrow consolidation.[50] In response to such actions, the citizens of countries applying the new model can use their computing resources together to establish competitive AI systems.

The even distribution of Synthetic Intelligence in society is also crucial in the context of the ever more widely discussed idea of Universal Basic Income (UBI).[51] In the case of current models of development and strengthening consolidation tendencies, granting UBI can be strictly tied to the obedience of an individual toward the entity that awards the income. This is in fact one of the examples that make an individual's rights conditional on their actions within the master–slave relationship. In the new model, UBI can be replaced with the income that every individual gains through the use of their personal computing power for their own gainful activity, which covers also providing the computing power to other entities. Optionally, UBI understood as a government-paid benefit could exist as an additional source of income. In this case, the funds for UBI would come from taxation of companies using society's computing power. The funds would go to the government's budget and then spread among citizens.

The even distribution of Synthetic Intelligence is also in line with the idea of social justice.[52] This model promotes equality of conditions by leveling the differences in intelligence potential as determined by evolution so far (which is still relatively low in comparison with possible differences in the Synthetic Intelligence potential). At the same time, the new approach does not determine the results which an individual can achieve by using their Synthetic Intelligence potential. The even distribution of Synthetic Intelligence among society can also strengthen an individual's mental stability in the AI era by increasing the key factors for the human psyche – a sense of security, stability and the awareness of real influence on surroundings and own future.[53]

3. Implementation of the new model – the starting point

The previous part covered the central concept of the model of even distribution of Synthetic Intelligence in society. To ensure that societies can function on the basis of this model, suitable implementation is needed. This is a topic that will be addressed in this part. The ideas presented here outline a sketch that is a starting point for further work. The premises are presented in points and grouped thematically.

Legally-established even distribution of Synthetic Intelligence

1. The foundation of the new approach is the even distribution of Synthetic Intelligence in society. This fundamental for the safety and stability of societies concept must be introduced to existing laws of the countries which will implement the new model. Further step should be taken to ensure that even distribution principle is entered into international treaties on human rights.

Physical deconsolidation, local clusters, and personal computing power (PCP)

1. The second fundamental aspect of the new approach is physical (geographic) deconsolidation of the computing power infrastructure and locating it close to the citizens who have been assigned this resources. In practice, the physical deconsolidation means that the computing power is located in fairly small local computing centers, i.e. local clusters. This approach is also intended to decrease the risk of an entity taking over or damaging a large portion of the computing power infrastructure that is located in a small number of places relatively far from the citizens.

2. People around a cluster (physically inhabiting a given area or spending most time there) become assigned to it, permanently obtaining own compute resources called as Personal Computing Power (PCP).
3. The size of local clusters (the number of people assigned to the cluster and its total computing power) should be a golden middle between safety, economy, and convenient use of the PCP. For example, if clusters are too small, they can be more difficult to maintain and ensure high security standards. If they are too large, there is a risk of excessive concentration of computing power in one place. It is worth considering a minimum and maximum cap, e.g. between 10,000 and 1,000,000 people assigned to a cluster. Another solution would be creating clusters along administrative borders, e.g. one cluster per district, city, or municipality.
4. Each person's PCP should increase evenly with the technological progress of next generations CPUs and increasing amount of global computing power.
5. Apart from the computing power infrastructure, each cluster should be also equipped with the necessary basic operating infrastructure to enable the cluster to function independently. This infrastructure should include, among others, communication modules (wire-based and wireless), autonomous power supply systems (e.g. solar-based, local small modular reactors, biofuel), necessary memory resources, software.
6. Local clusters must provide reliable and efficient communication with the surrounding community and the rest of the world (including other clusters).
7. The cluster's communication with the rest of the world must be able to be quickly curtailed if necessary through a physical procedure. In critical cases (e.g., remote takeover attempts by some entity), it should be possible to cut off communications completely.
8. To a large extent, the ongoing operation of the cluster should be the responsibility of the community around it and using its computing power.
9. At the same time, the long-term cluster maintenance and upgrade of its infrastructure could be done by external private or public entities. Nevertheless, any such work should be overseen by a team of people associated with the local cluster.
10. In case of a long-term change of the physical residence of persons, it should be possible to assign them to another cluster in order to increase safety and efficiency of PCP use.

Access to PCP

1. In order to use the computing power, two-factor authentication should be required (e.g. PIN/password and authenticating app/dongle).[54] At the same time, it would be recommended to use multi-factor authentication, e.g. with additional biometric data and behavior patterns.[55]
2. In the coming years, it will be necessary to ensure additional physical security measures. Particularly important will be the placement of physical security features in the processor chips, preventing unauthorized use of them.[56]
3. The resilience of the security measures implemented in software and hardware should continuously increase with the growth of global computing power potential.
4. The owner of a PCP can use his computing resources for any purpose that does not violate the rights and freedoms of others. This approach is a continuation of the existing way of free societies functioning where any person can use his brain's intellectual resources to perform any activity that does not violate the rights and freedoms of others.

PCP lease

1. PCP can be leased to other entities. The lessee can use PCP of other people to perform their own activities once digital agreements are signed and on condition that the purposes for which PCP is used are open. Current forms of employment and civil law contracts are intended to absorb a person's biological intelligence for specific tasks. PCP lease will work analogously, but in this case the tasks will be performed thanks to non-biological computing power coming from the local cluster.
2. A PCP lessee can be another physical person, a company, or a public institution.
3. An individual can lease a part of their PCP to one or many entities.
4. The lessor and lessee will conclude an agreement that will be satisfactory to both parties and that will bind them with clearly defined rights and obligations.
5. The PCP lease can cease if the applicable law, terms and conditions, or the agreement is breached or by termination of the agreement by one party.
6. Managing the process of validating lease agreements, checking their validity in light of changes in the law, and checking the consistency of declared and real usage goals of the PCP should be possible through open-source software that ensures transparency of the rules of its operation.
7. Lease agreements could be stored digitally in a distributed peer-to-peer (P2P) network based on smart contract technologies.[57]
8. In addition, it should be considered if the entities should be able to:
   a) lease PCP from any number of clusters,
   b) lease PCP from clusters from any area,
   c) lease any amount of PCP per cluster,
   d) lease any amount of global PCP.

Intellectual property – gathered data, know-how

1. The regulations on the data and know-how obtained by an entity remains unchanged. Legally gained skills, data, and knowledge must remain the property of the persons who created and gathered them on the same legal basis that currently functioning in democratic countries.
2. Each entity (physical persons, companies, public bodies) can generate and gather any amount of data, know-how and skills that it deems necessary and any amount of storage media that are necessary to keep them.
3. With each passing year, more and more data will be generated with non-biological computing power. Therefore, it is necessary to ensure that these processes are highly protected. In particular, this means ensuring:
   a) exclusive access of the owner to the non-leased PCP,
   b) exclusive access of the lessee to the leased PCP throughout the whole lease period,
   c) end-to-end (E2E) encryption of the communication between the owner's and lessee's devices and the local cluster that processes the data,[58]
   d) encryption of the communication between rest of clusters, sensors, external data sources.

Personal storage

1. In order to improve safety as well as performance, it is reasonable for the local clusters to contain also basic storage space for data of persons assigned to cluster – personal storage (PS). Similarly to PCP, access to personal storage should be an unalienable right of every person. Users should can use PS for any purposes they wish, e.g. to store application data, create a backup copy, or a vault for confidential data. The aim of PS is to increase the security of data storage and management in comparison with the situation when it is stored in the clusters

managed by third parties and located in distant places. Such approach can also influence performance. The data that is stored in a local cluster relatively close to the owners can be transferred quicker, without overloading the global network infrastructure.
2. Apart from E2E encryption, PS should also provide client-side encryption.[59]
3. The amount of the PS space should increase over time with the increasing PCP per citizen.

Computing power of devices beyond local clusters

1. Today's digital devices have specified computing power. These include, among others, house appliances, TVs, smart-home and smart-city devices, smartphones, notebooks, PCs, means of transport (air, sea, and land), machinery and robots, both office and industrial ones. Given the trend of equipping the next generation of devices with an increasing amount of computing power and the rapid increase of sophistication of AI technologies, it is crucial to adopt an appropriate strategy to ensure the security of their use of these devices in the external environment (i.e. outside the clusters).
2. The digital devices should be created in a way that their computing power is always applied to the intended use as specified by the manufacturer. The security measures that prevent the user from changing the intended use of the computing power should be implemented both on the software and hardware levels.
3. The digital devices that are designed for specific purposes should not be able to re-program themselves autonomically once they enter operation. In turn, their updates should be performed only by certified entities (manufacturers) after approval of such action by users.
4. All kinds of digital devices should be classified as one of two groups that defines their area of use:
    - local – stationary or with a small operating area around a local cluster. This group should include household appliances, audio-video devices, smart-home and smart-city devices, domestic and office robots, PCs, machinery and robots in industrial plants.
    - global – highly mobile, operating on an area that spans beyond a local cluster, including even a global area of operation. This group should include smartphones, notebooks, air, sea, and land means of transport.
5. The computing power in digital devices should have upper limits. This approach aims at limiting the risk of use of the digital devices by entities which could consolidate their large numbers and use their computing power for socially harmful purposes. The limits of permissible built-in computing power can be different for local and global devices.
6. For local devices, the permissible built-in computing power should be relatively low, but sufficient enough for independent operation of devices such as household appliances, audio-video devices, or smart-home systems which are moderately "intelligent".
7. In case of local devices with medium computing power requirements (e.g. advanced smart-home systems, domestic or office robots which are fairly "intelligent"), their built-in computing power should be purposefully insufficient for their autonomous operation. In this case, the higher functions having greater requirements should depend on the user's PCP in a local cluster.
8. In case of local devices with high computing power requirements (e.g. complex smart-city systems or advanced industrial robots), the built-in computing power should be purposefully insufficient for their autonomous operation. In this case, their operation should depend on the PCP of an adequately large group of people, e.g. co-owners, workers, or customers of the entity owning such a device.

9. In the case of global devices, the permissible built-in computing power could be higher than in local devices. These devices can have difficult access to local clusters' PCP and this could negatively impact the stability, performance, and safety of their operation. The permissible built-in computing power of global devices could be different for specific types of devices.
10. In the case of small global devices such as smartphones and notebooks, the allowable embedded computing power should be relatively small. If more resources are needed (for example, while using AI-based personal assistants), the user of such a device should use PCP resources located in the cluster.
11. For large and highly complex global devices (e.g. air, sea, and land means of transport or spaceships), the built-in computing power could be adequately larger to ensure their proper operation even if they are far away from clusters. At the same time, special safety measures should be adapted for these devices to avoid constructing fully autonomous machines with LLMs, which could obtain the Artificial General Intelligence that is similar to human intelligence.
12. The amount of the built-in computing power for the majority of global devices (with the exception of spaceships with a large interplanetary range) could become lower over time as the world will be covered by a high bandwidth network infrastructure, allowing effective communication with clusters located even at long distances from these devices.[60]
13. Individuals, companies, and public entities can accumulate any number of digital devices with relatively low computing power while keeping in mind that the operation of more 'intelligent' functions may require cluster resources.
14. Individuals can own global digital devices (especially means of transport) with significant embedded computing power in exchange for giving up an equivalent amount of their own (non-leased) PCP resources.

    Companies and public entities may own global digital devices (especially means of transportation) with significant embedded computing power in exchange for giving up an equivalent amount of their own (non-leased) PCP resources by people who have direct influence on the entity's operations or the beneficiaries of its operations (co-owners, management, employees, customers).

The role of technological companies and public bodies

1. Companies and public institutions do not own the computing power infrastructure. Such approach ensures two-sided dependence between the entities and society. Instead of creating the society's risk, mistrust, or dislike toward companies and governments, the new approach is based on maintaining mutual dependence, cooperation, and transparency.
2. Under the new model, hardware and software companies are, on the one hand, lessees (users) of the society's computing power which is necessary for operation, development, and competition on the free market. On the other hand, alignment with the society's expectations allow companies generate profit and provide new hardware or software technologies for the humanity's needs.
3. Hardware companies will still play the key role in maintaining the current IT technologies and in developing new ones to deliver them to communities gathered around local clusters. The manufacturers of CPUs would create their new generations and deliver them to upgrade, maintain, and expand cluster infrastructure. These companies would create also the CPUs for use in local and global devices beyond the clusters (in vehicles, smart-home systems, robots, and industrial machinery).

4. Software development companies remain responsible for developing current and developing new solutions. Unlike in the current model, these apps would not be hosted in a private infrastructure, but delivered directly to local clusters belonging to the society. Software companies could still develop "traditional" apps as well as AI solutions based on large language models (LLMs). In the latter case, it is necessary to create a comprehensive certification process for the created software to ensure their safety. Once certified, the LLMs would be implemented in local clusters.
5. Thanks to the strong mutual dependence of IT entities and the whole society, all parties can actively participate in partner-based process of creating standards, laws, and certificates that increase the social value and safety of the developed hardware and software.

Recommendations for R&D of new LLMs and LLMs deployed in clusters

1. LLMs should be developed only by certified entities (private and public) that meet a set of standards to ensure a high level of security at the stages of their development and use by the public.
2. The responsibility for potential damage caused by LLMs should be carried by the entities that develop them. This will force much more attention to the security area and the implementation of appropriate safety measures (e.g., approaches such as: Reinforcement Learning from Human Feedback, Explainable AI, Interpretable AI). [61, 62, 63]
3. LLM deployment should only take place in certified clusters. This can be ensured through a procedure of required acceptance of the deployment by two parties, i.e. by the certified vendor and the certified local cluster in which the LLM will be deployed.
4. It is worth considering setting upper limits of computing power for LLMs running in clusters. The limits could vary depending on the type of LLM and its spectrum of application. Technically, such limits could be implemented at the P2P network protocol level and possible to modified only by a decision of significant part of all citizens.

Recommendations for R&D and manufacturing of cluster infrastructure

1. R&D on technologies used in clusters (e.g. CPUs, communication devices, power supply technologies) should be spread across multiple locations to reduce the risk of over-consolidation of know-how in one place.
2. Also, the production of components itself should be dispersed to reduce the risk of excessive consolidation of physical means of production in one place.
3. Complete trace records should be kept of the life cycle of components used in clusters, as well as of devices outside clusters that have significant in-build computing power - especially global devices. Records should include distribution chains, inventory, assignment of components and finished products to specific entities, repair and utilization processes.

Transition to the new model

1. The transition to the new model should be a well-planned and gradually implemented process that is transparent for all parties. The transition should be spread over many months and performed without detriment to any entity that takes care of the safety and stability of society.
2. Gradual transition ensures that the total computing power used in a given country will not become lower during moving to the new model. This will ensure that private and public entities which use computing power can operate without interruption.

3. Entities that have significant computing power infrastructure should be able to sell it to local or central government bodies. This infrastructure should then be physically placed in created local clusters.
4. The role of owners and computing power services providers will be gradually phased out for tech companies as computing power infrastructure moves to local clusters. At the same time, the know-how and manufacturing potential of these entities will play the essential role in the continuous development and delivery of software and hardware (including processors) technologies used in the local clusters.
5. The establishment of local clusters should be funded by local and/or central government budgets. If needed, it can be also financed through repayable and/or non-repayable grants from other countries which have already implemented new model.
6. Financing of maintenance and development of local clusters should, over time, take place as much as possible from local budgets. The computing power potential of a given cluster should increasingly 'earn itself' - similarly as today the high level of 'intellectual human capital' of an individual and companies allows to sustain and grow themselves.

Global propagation of the new model

1. PCP should be evenly distributed among all the citizens of an international community that implements the new model of even distribution of Synthetic Intelligence.
2. The priority for the communities that implement the new model should be promoting it on a global scale to build a stable international safety architecture. To achieve this, the architecture must have a highly inclusive and partner-like nature.
3. Each country that wants to move onto the new model and become part of the new safety architecture must have this possibility once it introduces fundamental legal and social changes that guarantee the safety of its citizens.
4. Each country that meets this requirement should obtain the necessary support in the form technology transfer from other countries belonging to the new architecture. Poorer countries could also receive financial support (repayable or non-repayable) to effectively even out the Synthetic Intelligence potential in a given part of the world.
5. The goal of a highly inclusive policy should be to ensure that Synthetic Intelligence is evenly distributed throughout humanity and to equalize the many derived factors that can determine the level of civilizational development. Ultimately, it should lead to an architecture that ensures a high level of individual security and social stability across the globe.

4. Cooperation between US and China on the basis of the new model

It is highly desirable that the model of even distribution of Synthetic Intelligence be implemented internationally by the two major powers, i.e. the US and China, which are currently leading progress in the field of AI.[64] They have the largest know-how and computing power potential as well as significant influence over other countries through, among others, economic ties. The implementation of the new model by these entities may be important for the global de-escalation of the risk of consolidation of AI capabilities that lead to one-sided dependency, and for reducing the risk of catastrophe for all of humanity caused by a further "race to the precipice". The section below presents a sketch of the current state of relationship and major scenarios for the implementation of the new model on international stage. In the first scenario, both superpowers implement the new model. In the second scenario, only one superpower implements it. Importantly, either of the scenarios is more beneficial than the situation in which both countries continue the current high-risk models.

The current state of affairs between the global powers

1. China is afraid of the dynamic expansion of the US in the AI sector, as this may degrade its current status to a secondary entity on the international stage.
2. On the other hand, the US fears that if it does not continue to make fast progress in the field of artificial intelligence, China may gain a dominant position.
3. Both sides may also fear that the opposite side will develop AI in a less safe and socially beneficial way. This fear may paradoxically force further rapid development which, in the perception of each side, is seen as a better way out than surrendering the field to the others.
4. Both sides should become aware that this state of affairs is potentially suicidal for all the parties of the race driven by emotions, especially the fear of losing their position. Fear-fueled attitudes do not promote socially sustainable and cautious decision-making by both sides and have crucial impact on continuing the race to the precipice in the AI sector.
5. Despite the extremely risky current tendencies, both superpowers should realize that further AI development may lead to alternative, a win-win situation for both countries and the whole humanity thanks to the great potential that this technology holds.
6. Having this in mind, it is essential to create a common framework for actions that will be acceptable to both sides and that will ensure cooperation on de-escalating AI risks. It is possible to lay down a common ground for cooperation to have safe and sustainable AI development. The worst possible option is to assume that dialogue is pointless.
7. China and the U.S. should work toward the safest possible solution that is acceptable to all sides and that is better than incompatible values, extreme emotions, escalating tensions and a further race to the precipice.
8. It is crucial to agree on the future that both sides want to achieve. This will determine the goals, values, and actions that both sides will perceive as key ones and towards which they will be moving in the coming years.

Questions to be answered by both sides

The US should ask itself if it wants to risk the escalation of internal and global tensions. If the AI is consolidated by corporations and/or the government, the country could move onto an authoritarian path (though possibly with a facade of democracy), and further toward a totalitarian state. Sooner or later, the underprivileged part of society may become entirely dependent on the entities that consolidated the potential of Synthetic Intelligence, unless there is an existential catastrophe for all humanity caused by a race to the precipice before then. Is this what current Big Tech stakeholders, political elites, and all citizens of this country want for them and their close ones? Or rather they want a society of free people, where the potential of any individual is important and where they jointly decide about future.

China should ask itself if it wants to risk the escalation of internal and global tensions and to create a totalitarian state for its citizens that has never been seen before in the history of humanity (unless there is an existential catastrophe caused by a race to the precipice beforehand). If AI potential will be consolidated by government and related entities, the citizens sooner or later may be cease to be useful to this entity to maintain its status. Is this what the leadership in China wants for its civilization? Or maybe, they dream of a just society in line with the Chinese doctrines of a harmonious society and common prosperity.[65, 66] If China wants a harmonious society and common prosperity, its basis should not be consolidation of AI potential, but an even, harmonious distribution of intelligence in society.

Scenario 1: Cooperation between both sides on the basis of the new model

1. Both countries recognize the consolidation of AI potential as contrary to human rights and an existential risk to individual safety, social stability, and international order.
2. Both countries sign international agreements in which they declare development on genuine, safe for citizens, even distribution of Synthetic Intelligence.
3. The governments start implementation of new approach based on architecture of local clusters and unalienable personal computing power.
4. In order to ensure safe development of new approach, an international agency is established, which consists of representatives of all the countries that implement the new model including the US and China. The agency's responsibilities should include:

    I) Forming international teams that focus on: 1) monitoring; 2) record-keeping; 3) research and development; 4) auditing. Forming and reinforcing a culture of dialogue and cooperation between teams, governments, and citizens of the countries.

    II) Monitoring and following the agreements by state authorities to ensure even and unalienable access to PCP for all citizens. Monitoring the computing power of clusters to ensure even distribution of PCP for all citizens. Notifying about any irregularities, detecting their causes, and preventing their reoccurrence.

    III) Monitoring data on PCP lease by all private and public entities: amount of leased PCP, structure of leasing entities, and details on cluster locations providing PCP to entities. Notifying about any irregularities, detecting their causes, and preventing their reoccurrence.

    IV) Monitoring data on all LLMs deployed in clusters – computing power currently used by an LLM, cluster location providing PCP for a given LLM, current data on communication with other LLMs (which LLMs are communicated with and how intensively). Notifying about any irregularities, detecting their causes, and preventing their reoccurrence.

    V) Record keeping of life cycle for critical infrastructure components (especially CPUs), including research and development, manufacturing, distribution to clusters, relocation over time, disposal or recycling of used components. Recorded data should be kept in a public P2P network.

    VI) Record keeping of LLM life cycle in clusters, including information on its implementation, scaling, updates, and removal. Recorded data should be kept in a public P2P network.

    VII) Record keeping of key data on LLMs developed in laboratories: their purpose and social function, implemented safety measures, and details about entities and research teams. Recorded data should be kept in a public P2P network.

    VIII) Supporting research and development of key components for the cluster infrastructure. Coordinating actions of private entities and governmental bodies in those areas.

    IX) Determining joint technological stack, guidelines, and applicable standards for hardware and software (including LLMs) of the technologies used in clusters.

    X) Obligatory auditing of all the entities engaged in the development and maintenance of the critical components of the infrastructure and LLMs.

    The cooperation of international teams and the entire agency work is to ensure that the even distribution of Synthetic Intelligence is not just a declaration, but is implemented and functions in the safest and most socially sustainable way possible.

5. Collaboration among superpowers should constantly focus on building global stability based on the new model and de-escalating risks from AI and derived technologies.
6. Countries are also constantly promoting the new model to other countries through diplomatic efforts. Moreover, since the United States and China have considerable know-how in the creation of LLMs and are producing high-end elements of computing power infrastructure, the transition to the new model may become a highly desirable direction for authorities and citizens of other countries, resulting in the promotion of the new model on a global scale.

Scenario 2: One of the superpowers does not cooperate on the basis of the new model

1. The government of one of the two countires does not resign from the accumulation of Synthetic Intelligence potential.
2. In this scenario, the accumulation is done overtly or covertly. In the latter case, the country officially claims that it wants and is going to implement the new model (e.g. by creating local clusters) but it will do it by its own rules. In such case, there is a risk that in practice the access to computing power may be controlled by the government. If some citizens become disobedient, the government can limit or completely block the access to PCP. The government can also completely take over the citizens' access to PCP and use it for its own purposes, including those which may be potentially detrimental to society.
3. Despite the fact that one superpower is not using the new approach, the other superpower and other smaller countries can implement the model of evenly distributed synthetic intelligence, gradually increasing the safety of their citizens Even if only one superpower and smaller countries adopts the new approach it will still increase local and also global security. With pluralism and the open flow of information, society can have an open exchange of views, concepts, ideas, thereby being able to create the next generation of AI in a more efficient and socially responsible manner.
4. At the same time, countries operating under the new model should be careful with transfer outside the know-how and advanced technologies like CPUs, software (including advanced LLMs), and other critical infrastructure components that could help an authoritarian or totalitarian countries in building advantages over its citizens and other countries.
5. An exception may be the transfer of know-how, which increases the security of potentially risky AI systems. The purpose of such a transfer should be to reduce the risk of a global catastrophe that may be caused by a country's failure to implement appropriate safety standards as it continues to consolidate.
6. In the face of the lack of cooperation and non-zero risk of establishing a global authoritarian or totalitarian rule, the associated countries can counteract by creating their own competitive AI systems. Although the potential of computing power in the new model is decentralized, it allows PCP leasing and thus the building of powerful LLMs that can benefit from the significant resources of computing power leased from the public. This allows for the development of appropriately competitive AI using significant amounts of computing power in the face of consolidation trends on the other side and the risk of imposing power.
7. By following a new model based on the free flow of information, dialogue and cooperation, countries can win the global competition with totalitarianism. Gaining advantages can lead at a certain stage to a return of both sides to talks aimed at initiating cooperation on the basis of a socially sustainable model of even distribution of synthetic intelligence. To make this happen as early as possible, the associated countries must conduct continuous diplomatic efforts to turn the other side back from the trend of consolidation and start cooperation based on a new model.

Scenario 3: Both countries stay with current models and continue consolidation

In addition to the above two scenarios, it is also possible that the both superpowers will continue their consolidation tendencies, leading to a one-sided dependency relationship, a continuation of the race to the precipice, a further global decline of security and a risky escalation of tensions. Obviously, such a scenario is highly detrimental not only to the security of the underprivileged part of humanity, but to everyone, potentially leading to an event that risks the survival of the entire species. In such a situation, without following the alternative to the current trends of consolidation, we accept here and now a future without freedom for all humans, and even without humanity.

5. The potential impact of the new model on the de-escalation of AI risks

Possible influence on de-escalating risks: regional/global consolidation of Synthetic Intelligence leading to a one-way dependency relationship.

The implementation of the new model may reduce regional and global risks of concentrating synthetic intelligence capabilities (depending on whether one or both powers adopt the new model) including de-escalation of following threats:

- Transitioning the social order towards totalitarianism and developing a highly risky one-sided relationship of dependency for the underprivileged part of society,
- Discrepancy of the goals and values between the AI potential owner and society as a result of growing status differences,
- Decrease security of the underprivileged parts of society, degradation of their rights including even the right to continued existence.

Assuming that both superpowers, i.e. the US and China, jointly follow the new model in the coming years, the risk of consolidation can be significantly de-escalated on a global scale. These countries have the dominant position in terms of AI development, and production of computing power infrastructure and a strong economic impact on other developed and developing countries. As a result, the risk of consolidation of Synthetic Intelligence and power by any government and/or corporation can be effectively minimized. On the other hand, if only one of the two superpowers decides to implement the new model, the risk will be lowered on a regional scale at first, meaning within the country itself, economically related countries and in other countries that will follow along the new path. At the same time, the inclusive policy of the new model promotes its implementation on a global scale, which can further minimize the global risks.

Possible impact on de-escalating risks: regional/global "race to the precipice" leading to a catastrophe for all humanity

If one or both superpowers adopt the new model, its implementation may de-escalate the regional and global risks of the race to the precipice. This includes reducing the following threats:

- Loss of control over powerful AI that can lead degradation of *Homo sapiens* status or annihilation of entire species,
- Catastrophic incident caused by the misuse of AI or derivative technologies by humans,
- Common issues of mental nature in society.

If both superpowers follow the new model, common fear and mistrust can be minimized, stopping a further race to the precipice. Retreat from the path of consolidation of Synthetic Intelligence potential by both superpowers could lead to de-escalation of global tensions. Building trust, cooperation, increasing ethical, social, and technological standards raise the safety of AI technologies. On the other hand, if only one of the superpowers decides to follow a new model based on an even distribution of Synthetic Intelligence, this will obviously be less beneficial for global security. Nevertheless, if even one of superpower implements the new model, this will have a valuable impact on de-escalating the race to the precipice. With the cooperation of the free individuals of this society based on the free flow of information, it will be possible to develop and maintain high-quality ethical, social and technological standards, which may significantly reduce the scale of risks from AI direction.

The importance of the alternative

Given the current consolidation trends, it is important to build an alternative model based on an even distribution of Synthetic Intelligence potential within the entire society. It is possible that, in the coming years, part of countries will still continue the path of consolidation . In such a situation, on one side we will have countries in which the consolidation of power by companies and/or governments will become stronger and stronger, and the formation of unilateral dependence becomes more closer. Moreover, the potential for the creation, propagation and implementation of new ideas by the citizens of these countries will be increasingly suppressed or marginalized. On the other hand, we will have countries in which there is a mutual dependence relationship between companies/governments and society. Countries in which the flow and propagation of ideas cannot be suppressed or marginalized. With these key factors for both security but innovation, as well as the motivation to constantly build and improve this alternative model, free societies can achieve a civilizational advantage over totalitarian states, gradually leading to the spread of the new model on a global scale. By moving to a model of even distribution of Synthetic Intelligence, new countries will be able to count on the support of countries already operating in the new model. The benefit for both sides will be building a new, stable architecture that ensures the security of all individuals in the AI era. By following the current models towards ongoing consolidation, we, as free societies, are surrendering today, consenting to the further decline of individual security with each passing month, the further erosion of the social order and the escalation of existential risks for humanity. By following a new model based on the even distribution of Synthetic Intelligence, we are giving ourselves a real chance to create a safe and stable future for humanity, which, if current trends continue, may be irreversibly lost.

[1] The term “Synthetic Intelligence” is used interchangeably with the term “Artificial Intelligence”.

[2] LeCun, Yann, Yoshua Bengio, and Geoffrey E. Hinton. “Deep Learning.” Nature 521, no. 7553 (May 27, 2015): 436–44. https://doi.org/10.1038/nature14539.

[3] Wei, Jason Zhanshun, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, et al. “Emergent Abilities of Large Language Models.” ArXiv (Cornell University), June 15, 2022. https://doi.org/10.48550/arxiv.2206.07682.

[4] LLM development milestones:
https://timelines.issarice.com/wiki/Timeline_of_large_language_models.

[5] Kurzweil, Ray. The Singularity Is Near: When Humans Transcend Biology. Penguin Paperbacks, 2006.

[6] Lee, Kai-Fu, and Chen Qiufan. AI 2041: Ten Visions for Our Future. Random House, 2021.

[7] Bostrom, Nick. Superintelligence: Paths, Dangers, Strategies. Oxford University Press, US, 2014.

[8] Turner, Alexander J., Logan Thomas Smith, Rohin Shah, Andrew Critch, and Prasad Tadepalli. “Optimal Policies Tend to Seek Power.” *Cornell University*, December 3, 2019. https://doi.org/10.48550/arxiv.1912.01683.

[9] Public statement and open letter about AI Risks:
https://www.safe.ai/statement-on-ai-risk and: https://futureoflife.org/open-letter/pause-giant-ai-experiments/ (see also: https://time.com/6273743/thinking-that-could-doom-us-with-ai/).

[10] Estimation of the current and future number of AI safety researchers globally:

McAleese, Stephen. “Estimating the Current and Future Number of AI Safety Researchers.” LessWrong, September 28, 2022.
https://www.lesswrong.com/posts/mC3oeq62DWeqxiNBx/estimating-the-current-and-future-number-of-ai-safety.

[11] Center for Security and Emerging Technology. “Exploring Clusters of Research in Three Areas of AI Safety - Center for Security and Emerging Technology,” June 9, 2023.
https://cset.georgetown.edu/publication/exploring-clusters-of-research-in-three-areas-of-ai-safety/.

[12] Tellis, Ashley J., Alison Szalwinski, and Michael Wills. STRATEGIC ASIA: U.S.-China Competition for Global Influence, 2020.

[13] Schneider-Petsinger, Marianne, Jue Wang, Jie Yu, and James Crabtree. US-China Strategic Competition: The Quest for Global Technological Leadership, 2019.

[14] Ernst, Dieter. “Competing in Artificial Intelligence Chips: China’s Challenge amid Technology War,” March 30, 2020. https://ssrn.com/abstract=3564669.

[15] OpenAI statement and initiative highlights the importance of compute and talented teams. As OpenAI key developers stated: “We’ve experienced firsthand that the most dramatic AI systems use the most computational power in addition to algorithmic innovations, and decided to scale much faster than we’d planned when starting OpenAI”. More information: https://openai.com/blog/openai-lp.

[16] Roth, Emma. “Microsoft Spent Hundreds of Millions of Dollars on a ChatGPT Supercomputer.” The Verge, March 13, 2023. https://www.theverge.com/2023/3/13/23637675/microsoft-chatgpt-bing-millions-dollars-supercomputer-openai.

[17] Lee, Kevin Sengupta, Shubho “Introducing the AI Research Super Cluster — Meta’s cutting-edge AI supercomputer for AI research”, Meta AI Blog, January 24, 2022: https://ai.facebook.com/blog/ai-rsc/.

[18] Armstrong, Stuart D., Nick Bostrom, and Carl Shulman. “Racing to the Precipice: A Model of Artificial Intelligence Development.” AI & Society 31, no. 2 (May 1, 2016): 201–6. https://doi.org/10.1007/s00146-015-0590-y.

[19] Hogarth, Ian. “We Must Slow down the Race to God-like AI.” Financial Times, April 13, 2023. https://www.ft.com/content/03895dc4-a3b7-481e-95cc-336a524f2ac2.

[20] Russell, Stuart. Human Compatible: Artificial Intelligence and the Problem of Control. Penguin, 2019.

[21] Christian, Brian. “The Alignment Problem: Machine Learning and Human Values.” Perspectives on Science and Christian Faith 73, no. 4 (December 1, 2021): 245–47. https://doi.org/10.56315/pscf12-21christian.

[22] Karnofsky, Holden. “AI Could Defeat All Of Us Combined.” Cold Takes, September 1, 2022. https://www.cold-takes.com/ai-could-defeat-all-of-us-combined/.

[23] Urbina, Fabio L., Filippa Lentzos, Cédric F Invernizzi, and Sean Ekins. “Dual Use of Artificial-Intelligence-Powered Drug Discovery.” Nature Machine Intelligence 4, no. 3 (March 7, 2022): 189–91. https://doi.org/10.1038/s42256-022-00465-9.

[24] Robert A. Freitas, Jr, “Molecular manufacturing: Too dangerous to allow?”, Nanotechnology Perceptions no. 2 (2006) 15–24. http://www.rfreitas.com/Nano/MMDangerous.pdf.

[25] Kırdemir, Can Kasapoğlu Barış. “Artificial Intelligence and the Future of Conflict - New Perspectives on Shared Security: NATO’s Next 70 Years.” Carnegie Europe, November 28, 2019. https://carnegieeurope.eu/2019/11/28/artificial-intelligence-and-future-of-conflict-pub-80421.

[26] Wong, Yuna Huh, John Yurchak, Robert W. Button, Aaron B. Frank, Burgess Laird, Osonde A. Osoba, Randall Steeb, Benjamin N. Harris, and Sebastian Joon Bae, “Deterrence in the Age of Thinking Machines.” Santa Monica, CA: RAND Corporation, January 27, 2020. https://doi.org/10.7249/RR2797.

[27] Scott, David, Bart Valley, and Brooke A. Simecka. “Mental Health Concerns in the Digital Age.” International Journal of Mental Health and Addiction 15, no. 3 (June 17, 2016): 604–13. https://doi.org/10.1007/s11469-016-9684-0.

[28] More on the social impact of unsustainable technological progress that can be strongly caused by AI:
Harari, Yuval Noah. Homo Deus: A Brief History of Tomorrow. Random House, 2016.

[29] Webb, Amy. The Big Nine: How the Tech Titans and Their Thinking Machines Could Warp Humanity. Hachette UK, 2019.

[30] Zuboff, Shoshana. The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. PublicAffairs, 2018.

[31] Janna Anderson, Lee Rainie "Concerns about democracy in the digital age" Pew Research Center, February 21, 2020. https://www.pewresearch.org/internet/2020/02/21/concerns-about-democracy-in-the-digital-age/.

[32] AI Now Institute. "Tech and Financial Capital." AI Now Institute, April 11, 2023. https://ainowinstitute.org/spotlight/tech-and-financial-cap.

[33] Waters, Richard. "Big Tech Is Racing to Claim Its Share of the Generative AI Market." Financial Times, April 20, 2023. https://www.ft.com/content/42e3e384-e79c-41c2-ab36-82c35456e7c6.

[34] Datta, Nivedita Balu Tiyashi. "Tech Companies to Highlight AI in Earnings; Investors Focus on Profits." Reuters, April 24, 2023. https://www.reuters.com/article/tech-results-idCAKBN2WL0LM.

[35] Hoppner, Thomas, and Luke Streatfeild. "ChatGPT, Bard &Amp; Co.: An Introduction to AI for Competition and Regulatory Lawyers." Social Science Research Network, January 1, 2023. https://doi.org/10.2139/ssrn.4371681.

[36] Beraja, Martin, Andrew Kao, David J. Yang, and Noam Yuchtman. "AI-Tocracy," November 1, 2021. https://doi.org/10.3386/w29466.

[37] Nick Bostrom, 'The Vulnerable World Hypothesis' - Global Policy, 2019: Bostrom, Nick. "The Vulnerable World Hypothesis." Global Policy 10, no. 4 (September 6, 2019): 455–76. https://doi.org/10.1111/1758-5899.12718.

[38] Wang, Maya. "China's Techno-Authoritarianism Has Gone Global: Washington Needs to Offer an Alternative." Foreign Affairs, April 19, 2023. https://www.foreignaffairs.com/articles/china/2021-04-08/chinas-techno-authoritarianism-has-gone-global.

[39] Roberts, Huw, Josh Cowls, Jessica Morley, Mariarosaria Taddeo, Vincent M. Wang, and Luciano Floridi. "The Chinese Approach to Artificial Intelligence: An Analysis of Policy, Ethics, and Regulation." AI & SOCIETY 36, no. 1 (June 17, 2020): 59–77. https://doi.org/10.1007/s00146-020-00992-2.

[40] ChatGPT: https://chat.openai.com, Google Bard: https://bard.google.com.

[41] More on the influence of Big Tech corporation on societies through technology:
Zuboff, Shoshana. The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. PublicAffairs, 2018.
and:
"How Big Tech Is Reshaping Governance." The Consilience Project, December 19, 2022. https://consilienceproject.org/how-big-tech-is-reshaping-governance/.

[42] Max Bank, Felix Duffy, Verena Leyendecker, Margarida Silva. "The Lobby Network: Big Tech's Web of Influence in the EU," Corporate Europe Observatory and LobbyControl e.V. Brussels and Cologne, August 2021
https://corporateeurope.org/en/2021/08/lobby-network-big-techs-web-influence-eu.

[43] Emily Birnbaum, "Tech Spent Big on Lobbying Last Year," POLITICO January 24, 2022. https://www.politico.com/newsletters/morning-tech/2022/01/24/tech-spent-big-on-lobbying-last-year-00001144.

[44] De Sola Pool, Ithiel. Technologies of Freedom. Harvard University Press, 1983.

[45] Morrisett, Lloyd. "Habits of Mind and a New Technology of Freedom." First Monday, September 2, 1996. https://doi.org/10.5210/fm.v1i3.483.

[46] Turner, Fred. From Counterculture to Cyberculture: Stewart Brand, the Whole Earth Network, and the Rise of Digital Utopianism. University of Chicago Press, 2006.

[47] The Treaty on the Non-Proliferation of Nuclear Weapons (NPT), 1968. More info: https://www.iaea.org/topics/non-proliferation-treaty.

[48] Asilomar and Recombinant DNA, 1975. More info: https://www.nobelprize.org/prizes/chemistry/1980/berg/article/.

[49] Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on their Destruction, 1997. More info: https://www.opcw.org/chemical-weapons-convention.

[50] Lee, Kai-Fu. AI Superpowers: China, Silicon Valley, and the New World Order. Houghton Mifflin, 2018.

[51] Bidadanure, Juliana. "The Political Theory of Universal Basic Income." Annual Review of Political Science 22, no. 1 (May 11, 2019): 481–501. https://doi.org/10.1146/annurev-polisci-050317-070954.

[52] Rawls, John. A Theory of Justice: Original Edition. Cambridge, MA and London, England: Harvard University Press, 1971. https://doi.org/10.4159/9780674042605.

[53] Over time, the larger the amount of synthetic intelligence assigned to an individual, the lower the difference of the total intelligence (i.e. biological alongside synthetic) potential will become for each individual.

[54] More information about Two-Factor Authentication: https://www.ibm.com/topics/2fa.

[55] More information about Multi-Factor Authentication: https://www.ibm.com/topics/multi-factor-authentication.

[56] Hilson, Gary "Hardware-Based Security Emerges for AI." embedded.com, February 15, 2023. https://www.embedded.com/hardware-based-security-emerges-for-ai/.

[57] More information about Smart Contracts: IEEE Conference Publication | IEEE Xplore. "An Overview of Smart Contract: Architecture, Applications, and Future Trends," June 1, 2018. https://ieeexplore.ieee.org/abstract/document/8500488. See also: https://www.ibm.com/topics/smart-contracts.

[58] More information about E2E encryption: "A Deep Dive on End-to-End Encryption: How Do Public Key Encryption Systems Work?," Surveillance Self-Defense https://ssd.eff.org/module/deep-dive-end-end-encryption-how-do-public-key-encryption-systems-work.

[59] More information about client-side encryption: “Why Client-Side Encryption Is the Next Best Idea in Cloud-Based Data Security,” Information Systems Security https://web.archive.org/web/20190113021511/http://www.infosectoday.com/Articles/Client-Side_Encryption.htm.

[60] In particular, this may be possible with the technologies of fast wireless communication, e.g. 6G or Starlink.

[61] Christiano, Paul F., Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep Reinforcement Learning from Human Preferences. Neural Information Processing Systems. Vol. 30, 2017. https://doi.org/10.48550/arXiv.1706.03741.

[62] Gunning, David, Mark J. Stefik, Jaesik Choi, Timothy M. Miller, Simone Stumpf, and Guang-Zhong Yang. “XAI—Explainable Artificial Intelligence.” Science Robotics 4, no. 37 (December 18, 2019). https://doi.org/10.1126/scirobotics.aay7120.

[63] Meng, Kevin, David Bau, Alex Andonian, and Yonatan Belinkov. “Locating and Editing Factual Associations in GPT.” Advances in Neural Information Processing Systems 36 (February 10, 2022). https://doi.org/10.48550/arxiv.2202.05262.

[64] Miller, Chris. Chip War: The Fight for the World’s Most Critical Technology. Simon and Schuster, 2022.

[65] Guo, Sujian and Baogang Guo. China in Search of a Harmonious Society, 2008.

[66] Guoguang, Wu. “China’s Common Prosperity Program: Causes, Challenges, and Implications,” Asia Society, 2022. https://asiasociety.org/policy-institute/chinas-common-prosperity-program-causes-challenges-and-implications.